

Ten Years of Global Evolution of the Human Respiratory Syncytial Virus BA Genotype with a 60-Nucleotide Duplication in the G Protein Gene^{▽†}

Alfonsina Trento,^{1,4*} Inmaculada Casas,² Ana Calderón,² Maria L. Garcia-Garcia,³ Cristina Calvo,³ Pilar Perez-Breña,² and José A. Melero^{1,4}

Unidad de Biología Viral¹ and Unidad de Virus Respiratorios,² Centro Nacional de Microbiología, Instituto de Salud Carlos III, Majadahonda, 28220 Madrid, Spain; Pediatrics Department, Severo Ochoa Hospital, 28911 Madrid, Spain³; and Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERES), Instituto de Salud Carlos III, Majadahonda, 28220 Madrid, Spain⁴

Received 15 February 2010/Accepted 14 May 2010

The emergence of natural isolates of human respiratory syncytial virus group B (HRSV-B) with a 60-nucleotide (nt) duplication in the G protein gene in Buenos Aires, Argentina, in 1999 (A. Trento et al., *J. Gen. Virol.* 84:3115–3120, 2003) and their dissemination worldwide allowed us to use the duplicated segment as a natural tag to examine in detail the evolution of HRSV during propagation in its natural host. Viruses with the duplicated segment were all clustered in a new genotype, named BA (A. Trento et al., *J. Virol.* 80:975–984, 2006). To obtain information about the prevalence of these viruses in Spain, we tested for the presence of the duplicated segment in positive HRSV-B clinical samples collected at the Severo Ochoa Hospital (Madrid) during 12 consecutive epidemics (1996–1997 to 2007–2008). Viruses with the 60-nt duplication were found in 61 samples, with a high prevalence relative to the rest of B genotypes in the most recent seasons. Global phylogenetic and demographic analysis of all G sequences containing the duplication, collected across five continents up until April 2009, revealed that the prevalence of the BA genotype increased gradually until 2004–2005, despite its rapid dissemination worldwide. After that date and coinciding with a bottleneck effect on the population size, a relatively new BA lineage (BA-IV) replaced all other group B viruses, suggesting further adaptation of the BA genotype to its natural host.

Human respiratory syncytial virus (HRSV), a member of the Pneumovirus genus within the Paramyxoviridae family, is recognized as the leading agent responsible for severe respiratory infections in the pediatric population (31, 34, 35) and a pathogen of considerable importance in vulnerable adults (23, 24). The global respiratory syncytial virus (RSV) disease burden is estimated at 64 million cases and 160,000 deaths every year (70). This virus causes regular seasonal epidemics which take place during the winter months in temperate countries or during the rainy season in tropical areas (12). A peculiar aspect of HRSV is that the immune response produced by infection does not confer long-lasting protection, which is why reinfections are common throughout life (30).

Neutralization tests performed with hyperimmune serum (16) and reactivity with specific monoclonal antibodies (4, 45) were used to classify HRSV isolates into two antigenic groups, A and B, which correlated with genetically distinct viruses (18). The main differences between these two groups are located in the major attachment G protein. This protein is a type II glycoprotein that shares neither sequence nor structural features with the attachment proteins (HN or H) of other

paramyxoviruses (69), and it represents one of the targets of the immune response (27, 43). The full-length membrane-bound G protein (Gm) of 292 to 319 amino acids (depending on the viral strain) is also expressed in a secreted version (Gs) that lacks the transmembrane domain due to alternative initiation of translation at a second in-frame AUG codon in the G open reading frame (M48) (52). The G protein is the viral gene product with the highest degree of antigenic and genetic diversity among viral isolates (4, 18, 28, 45). Most changes are concentrated in two hypervariable regions that flank a highly conserved central region of the G protein ectodomain, which includes a cluster of four cysteines and the putative receptor binding site (43). It has been suggested that antigenic differences within this protein could facilitate repeated HRSV infections (37, 59). In addition, positive selection of amino acid changes was observed in the two hypervariable regions of the G protein ectodomain (7, 43, 71, 73, 74). One of the hypervariable regions, located in the C-terminal one-third of the G molecule, contains multiple epitopes recognized by monoclonal antibodies (43), suggesting that immune selection of new variants by antibodies may contribute to generation of HRSV diversity.

Phylogenetic studies based on sequence analysis of the G protein have identified numerous genotypes in the antigenic groups A and B that show complex circulation patterns, since multiple genotypes of both antigenic groups may circulate within the same season and community, with one or two dominant genotypes being replaced in successive years (13, 14, 26, 27, 32, 49, 50). Each community shows a seasonal circulation

* Corresponding author. Mailing address: Centro Nacional de Microbiología, Instituto de Salud Carlos III, Majadahonda, 28220 Madrid, Spain. Phone: 34 91 8223929. Fax: 34 91 5097919. E-mail: altrento@isciii.es.

† Supplemental material for this article may be found at <http://jvi.asm.org/>.

▽ Published ahead of print on 26 May 2010.

pattern of genotypes, probably determined by local factors, such as the level of herd immunity to certain strains (3, 14, 49).

The capacity of the G protein to accommodate drastic sequence changes was illustrated best by three antigenic group B viruses isolated in Buenos Aires, Argentina, in 1999 that contained a duplication of 60 nucleotides (nt) in the C-terminal third of the G protein gene (63). The global dissemination of these viruses allowed us to use the duplicated segment as a natural tag to reexamine the evolution of HRSV during propagation in its natural host. Phylogenetic analysis of G sequences revealed that all viruses with the duplicated segment clustered in a new genotype, named BA, and this finding supported the idea of a common ancestor for all viruses with the 60-nt duplication, dated about 1998 (64). The limited information about the molecular epidemiology of HRSV in Spain, together with an increase in G sequences with the duplicated segment reported worldwide, prompted us to conduct both a local search in Madrid for these viruses and a global phylogenetic analysis of HRSV with the 60-nt duplication from the time that these viruses were first detected, taking into account the geographic and temporal distribution of each isolate.

MATERIALS AND METHODS

Clinical samples. Nasopharyngeal aspirates were taken in phosphate-buffered saline from children hospitalized in the Severo Ochoa Hospital (Madrid, Spain) during 12 consecutive seasons (1996-1997 to 2007-2008). Respiratory samples were sent to the laboratory of respiratory viruses at the Centro Nacional de Microbiología (Instituto de Salud Carlos III, Majadahonda, Madrid, Spain). HRSV detection and classification into groups A and B were performed by multiplex nested reverse transcriptase PCR (RT-PCR) as described by Coiras et al. (17). Samples were stored at -80°C until further analysis. Virus nomenclature uses a letter code representing the site of isolation (e.g., "MAD" represents Madrid) followed by the isolate number and year of isolation for samples from the Southern Hemisphere or the first year of the epidemic season for samples from the Northern Hemisphere.

RNA extraction, DNA amplification, and sequencing of group B G sequences. For screening purposes, the presence of the 60-nt duplication in the G protein gene of HRSV group B samples was detected by a real-time RT-PCR as follows. Total RNA was extracted directly from frozen clinical specimens by using the RNA QIAamp MiniElute virus spin kit and the automated system QIAcube (Qiagen) according to the manufacturer's instructions. cDNA was obtained using 5 μl of total RNA in a mix containing 65 ng of the oligonucleotide BG9(-) (5'-GGAATTCGTCGACTTTTTTTTGAATAA-3'; negative sense), 0.5 mM (each) deoxynucleoside triphosphates (Promega), 10 U of avian myeloblastosis virus (AMV) reverse transcriptase, and reaction buffer (Roche) in a final volume of 10.8 μl . The reaction mix was incubated for 30 min at 42°C , followed by 15 min at 95°C . Next, the total volume of the RT reaction mixture was incubated with a mix of 65 ng of the oligonucleotide OG715(+) (5'-AAACCAACCCCAAGA CCACA-3'; nucleotides 715 to 735 of the G protein gene from the CH18537 strain; positive sense), 65 ng of the oligonucleotide OG813(-) (5'-GAGGGAT TGCTGTTGGATTGT-3'; nucleotides 715 to 735 of the G protein gene from the CH18537 strain; negative sense), and 15 μl of Power SYBER green PCR master mix (Applied Biosystems) in a final volume of 10.8 μl . Thermocycling was performed on a StepOne real-time PCR instrument (Applied-Biosystems) under the following conditions: 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 65°C for 1 min. After amplification, melting-curve analysis was performed with the following steps: denaturation at 95°C for 1 min, annealing at 60°C for 1 min, and 100 cycles of 0.3°C increments (10 s each). cDNAs of strains CH18537 and BA/3833/99 were used as controls for the absence and presence of duplication, respectively. Finally, a fragment of the last 417 nt of the G protein gene with the duplicated segment was amplified as previously described (64), except that BG9(-) was used as an antisense oligonucleotide. Ten microliters of the amplified DNAs were resolved by electrophoresis in 2% agarose gels, stained with ethidium bromide, and visualized with UV light. PCR products were then subjected to forward and reverse cycle sequencing with the BigDye terminator sequencing kit (Applied Biosystems) and the PCR primer sets, as described previously (64).

Sequence data and nucleotide and amino acid sequence analysis. A total of 347 partial sequences (330 nt long) of HRSV group B viruses with the 60-nt duplication in the G protein gene isolated worldwide were retrieved from GenBank for comparative studies. Of these, 187 sequences were unique for each country and contained a minimum length of 330 nt (detailed in Table S1 in the supplemental material). These sequences were aligned with ClustalX v1.81 software (62) and manually edited with the BioEdit software program, version 7.0.9 (33). The nucleotide and amino acid pairwise distance values (p-distances) were calculated using the MEGA software program, version 4.0 (61). The inferred amino acid sequences (using universal code) and polymorphisms were analyzed with DNAsp 4.0 software (53).

Phylogenetic analysis. Maximum-likelihood phylogenetic trees were obtained using the PAUP* software package, version 4.0b10 (60). The best-fitting nucleotide substitution model (general time reversible [GTR] + Γ_4) was determined from the data with the ModelTest software program, version 3.06 (51). For each tree, the reliability of phylogenetic grouping was determined through a bootstrap resampling analysis, using 1,000-replicate neighbor-joining trees under the maximum-likelihood (ML) substitution model. Trees were edited using the Treeview software program, version 1.5.2 (47), and the MEGA program, version 4.0 (61).

Evolutionary analysis. Rates of nucleotide substitution per site, the time of the most recent common ancestor (MRCA), and the key aspect of demographic history, particularly changes in the population size, quantified as $N_e\tau$, where N_e is the effective number of infections and τ is the generation time in years, were estimated using a Bayesian Markov chain Monte Carlo (MCMC) method available in the BEAST software package (<http://beast.bio.ed.ac.uk>) (19, 20). This method analyzes the distribution of branch lengths between viruses isolated at different times (year of collection) among millions of sampled trees. The HKY85 + Γ_4 substitution model was used, since more-complex models did not result in statistical convergence. The data set was analyzed using the Bayesian skyline model, assuming a relaxed (uncorrelated log normal) molecular clock. MCMC chains were run for sufficient time to achieve convergence (as assessed using the TRACER software program; <http://tree.bio.ed.ac.uk/download.html?name=tracer&version=v1.4.1&id=60&num=2>). Statistical uncertainty in parameter estimates is given by the 95% highest-probability density (HPD) values. The data obtained in the MCMC analysis were also used to infer a maximum clade credibility (MCC) tree using the software programs Tree-Annotator v1.4.7 (21) and FigTree v1.2.1 (<http://tree.bio.ed.ac.uk/software/figtree/>) for editing purposes. The MCC tree depicts the phylogenetic relationship against the time of sampling (year) of each isolate.

Nucleotide sequence accession numbers. The sequences from Madrid reported here were deposited in the GenBank database under accession numbers GQ150687 to GQ150743.

RESULTS

HRSV G gene sequences with a 60-nt duplication from samples collected in Madrid throughout 12 epidemic seasons. A total of 1,594 clinical samples collected from patients of the Severo Ochoa Hospital (Madrid, Spain) during 12 consecutive epidemic seasons (1996-1997 to 2007-2008) tested positive for HRSV and were classified in antigenic group A or B by RT-PCR of the F protein gene, as described by Coiras et al. (17). Both antigenic groups cocirculated over the 12 epidemics (Table 1). Although the overall prevalence of group A (54.7%) was slightly higher than that of group B (45.3%), group B viruses predominated in five epidemics (Table 1). A subset of 179 group B samples, representative of all epidemics, was selected for further analysis. A G gene segment, encompassing nucleotides 715 to 813, was amplified from those samples. Of the 179 amplicons, 61 showed an increase in the melting temperature and a slow-mobility band of the amplified product, indicative of a 60-nt duplication in the G protein gene (see Materials and Methods). The duplication was subsequently confirmed by DNA sequencing for 57 samples. Sequences could not be obtained in four cases due to scarce availability of the clinical specimen.

The first viruses with a 60-nt duplication in the G gene were detected in Madrid during the 1998-1999 epidemic, represent-

TABLE 1. Circulation of HRSV viruses in Madrid throughout 12 consecutive epidemics

Season of epidemic (yr)	No. of HRSV samples ^a	No. (%) of group B samples ^b	No. of studied samples ^c	No. (%) of samples with 60-nt duplication ^d	No. of samples with C-terminal sequences (472 nt) ^e
1996-1997	176	25 (14.2)	15	0 (0,0)	0
1997-1998	146	110 (75.3)	36	0 (0,0)	0
1998-1999	142	90 (63.4)	21	3 (14,3)	3
1999-2000	199	1 (0.5)	1	1 (100,0)	0
2000-2001	142	43 (30.3)	16	0 (0,0)	0
2001-2002	125	81 (64.8)	16	1 (6,3)	1
2002-2003	44	5 (11.4)	4	2 (50,0)	2
2003-2004	233	208 (89.3)	16	9 (56,3)	9
2004-2005	79	20 (25.3)	19	15 (78,9)	15
2005-2006	58	25 (43.1)	16	13 (81,3)	12
2006-2007	118	81 (68.6)	10	8 (80,0)	8
2007-2008	132	33 (25.0)	9	9 (100,0)	7

^a Number of samples that tested positive by multiplex RT nested PCR for HRSV.

^b Number of HRSV samples that were classified in antigenic group B by RT-PCR analysis of the F protein gene.

^c Number of HRSV group B samples selected for amplification of the G protein gene segment (nucleotides 715 to 813).

^d Number of HRSV group B samples that showed an amplified product with a $T_m > 80^\circ\text{C}$ and a slow-mobility DNA band in agarose gels, indicative of a 60-nt duplication in the G protein gene.

^e Number of group B samples with a 60-nt duplication for which a C-terminal G sequence could be obtained.

ing 14.3% of all group B samples from that year. Although the prevalence of HRSV strains with the duplication among group B viruses was low until 2001-2002 (but in 1999-2000, only one group B sample was available), this parameter increased after the 2002-2003 epidemic, reaching values close to 100% in recent years.

Interestingly, the first sequences with the 60-nt duplication found in this analysis (December 1998) were present in samples collected a few months earlier than the date of the first viruses with the duplication reported from Buenos Aires (June 1999) (63). However, the oldest sequence from the Madrid samples contained two nucleotide changes in the first duplicated region (see Fig. S1 in the supplemental material). Therefore, the ancestor of the viruses with the 60-nt duplication is expected to be more closely related to the viruses isolated in Buenos Aires in 1999, which contained two identical copies of the duplicated segment, than to Madrid viruses.

Phylogenetic relationship of Madrid sequences with the 60-nt duplication. Nucleotide sequences with the duplication in the C-terminal third of the G protein gene (nucleotides 510 to 981) could be obtained for 57 samples from Madrid (Table 1). Alignment of these sequences confirmed the presence in all of them of a duplicated segment of 60 nt, starting after position 792 (the number refers to the sequence of group B strain CH18537), coinciding with the duplication previously found in viruses from Buenos Aires (63, 64). Identical sequences were observed occasionally in different samples collected during the same epidemic season, except for MAD/4163/03, which was found in two consecutive epidemics (2003-2004 and 2004-2005).

The maximum-likelihood method was used to construct a phylogenetic tree of 30 unique partial G sequences from Madrid with the 60-nt duplication (Fig. 1). The sequences were clustered into four branches (MAD-I to MAD-IV). All the G sequences with the duplication from viruses that circulated until 2002-2003 were clustered in MAD-I. This branch was apparently extinguished after the 2004-2005 epidemic, as no further sequences that group in MAD-I were identified after

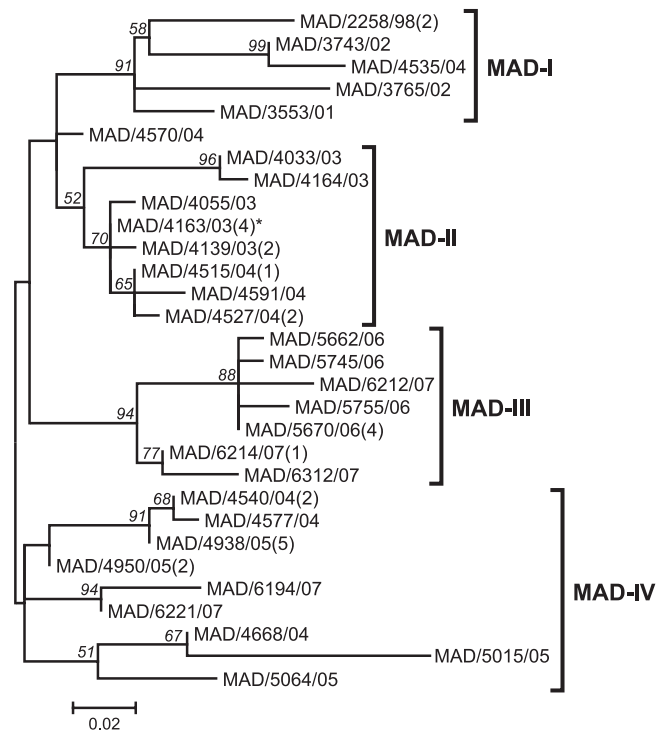


FIG. 1. Phylogenetic tree of HRSV-B sequences from Madrid with a 60-nt duplication. Unique partial G sequences (nucleotides 510 to 981) with a 60-nt duplication obtained from clinical samples collected in Madrid (Table 1) were used to construct a phylogenetic tree by the maximum-likelihood method (see Materials and Methods). The lengths of horizontal lines are proportional to the genetic distances between viruses. The bar represents 0.02 nucleotide substitutions per site, and the tree is unrooted. Numbers at the internal nodes represent the bootstrap probabilities (1,000 replicates). Only bootstrap values > 50 are shown. Branches are indicated at right by brackets. The number of MAD sequences identical to those shown in the figure is indicated between parenthesis at right of the sample name, and the asterisk indicates sequences identified in more than one season.

Year of isolation	Group and genotype	Brazil	Belgium	Madrid	Buenos Aires	New Delhi	South Africa	Niigata
1998	B		76/102	90/142	5/20			
	BA		0	3/21	0			
1999	B	23/97		1/199	55/94			
	BA	2/19		1/1	8/55			
2000	B	15/84		43/142	0/44			
	BA	0/0		0/0	0/0			
2001	B		73/86	81/125	10/56	1/1		
	BA		25/37	1/16	1/9	1/1		
2002	B	59/131		5/44	45/63	0/8		45/54
	BA	12/35		2/4	23/45	0		41/45
2003	B	46/123		208/233	24/92	2/22		16/99
	BA		10/10	9/16	12/24	2/2		16/16
2004	B		68/90	20/79	10/57	8/17		
	BA		15/15	15/19	6/10	8/8		
2005	B	UD	12/83	25/58				
	BA	90/90	3/3	13/16				
2006	B			81/118			7/13	
	BA			8/10			7/7	
2007	B			33/132				
	BA			9/9				

FIG. 2. HRSV group B and BA genotype prevalence in seven countries between 1998 and 2007. Epidemiologic data for Brazil (7), Belgium (73, 75), Madrid (this work), Buenos Aires (26, 64, 67), New Delhi (48), South Africa (68), and Niigata (55) are included in this study. Prevalence of group B over group A $\geq 50\%$ is highlighted in green. Prevalence of the BA genotype over other group B genotypes $\geq 50\%$ is highlighted in pink. B values, numbers of group B strains out of the total HRSV-positive samples. BA values, numbers of sequences with the 60-nt duplication out of the total group B sequences studied. The year of isolation corresponds to the first year of an epidemic season, i.e., 2001 for the 2001-2002 epidemic. UD, unpublished data.

this date. In 2003-2004 emerged a new branch, MAD II, which prevailed in the following epidemic (2004-2005), clustering 8 out of 15 sequences from 2004-2005. The remaining sequences from this epidemic clustered in the branch MAD-IV with sequences from samples isolated in later years, except for MAD/4570/04, which shared a phylogenetic relationship with branches MAD-I and -II. The branch MAD-III included all sequences of the epidemic season 2006-2007 and the majority of sequences from the following season (2007-2008). Branch MAD-IV was the most heterogeneous regarding the date of samples. This branch, with low bootstrap, contained both local viruses that acquired changes with time and viruses probably imported from other regions (see below).

Global distribution of HRSV-B G sequences with the 60-nt duplication over 10 years. Since the first global analysis of the BA genotype, published in 2006 (64), the number of reported G sequences with the 60-nt duplication has increased considerably. A total of 404 G sequences with the duplication (most of them partial sequences), including the 57 sequences from Madrid reported here, are available in GenBank (as of April 2009; see Table S1 in the supplemental material). Sequences with the duplication come from samples isolated in 12 countries throughout the last 10 years. The largest number of sequences with the duplication have been reported in Brazil (169 samples) (7; M. C. O. Souza, T. K. Matsumoto, L. R. A. Vaz-de-Lima, N. N. Sato, M. M. Salgado, H. I. Z. Requejo, M. A. Hong, M. L. Barbosa, C. A. F. Oliveira, P. C. Benelli, R. Y. Shimokawa, R. Pecchini, E. Berezin, S. P. Duarte, C. Schvartsman, E. L. Durigon, and M. Ueda, submitted for publication; J. L. Proenca-Modena, F. E. Paula, C. Ansarah-Sobrinho, A. M. Saranzo, A. E. Santos, M. A. Iwamoto, O. A. L. Cintra, and E. Arruda, submitted for publication), followed by Belgium (67 samples) (73, 75), Madrid (57 samples) (this

work), Buenos Aires (49 samples) (63, 64), and New Delhi (39 samples) (48; P. Bharaj, W. M. Sullender, H. S. Chahar, C. John, V. Tyagi, S. K. Kabra, and S. Broor, submitted for publication). Excluding the three samples isolated in Madrid in 1998, all samples were collected after the original report of viruses with the 60-nt duplication isolated in Buenos Aires in 1999.

Since epidemiologic information was available for seven places (including Madrid) (7, 26, 48, 55, 64, 67, 68, 75), the prevalences of both HRSV group B and the BA genotype were examined in these locations (Fig. 2). The proportion of viruses with the duplication increased over time, and the BA genotype now prevails over the rest of the group B genotypes (highlighted in pink in Fig. 2), irrespective of the prevalence of group B over group A in a given season (highlighted in green in Fig. 2). For example, in Belgium, 100% of the group B viruses found in the three latest epidemics (2003-2004 to 2005-2006) contained the duplication (75), and findings were similar in Brazil in 2005 (7), in New Delhi in 2001-2002 and 2004-2005 (48), and in South Africa in 2006 (68). The epidemiologic data published from Niigata (Japan) showed a high prevalence of the BA genotype among group B viruses (91.1% in 2002-2003 and 100% in 2003-2004); however, unfortunately, only 4 of the 57 G sequences with the duplication analyzed by Sato (55) are available in GenBank. Analyzed samples from Madrid showed the same tendency, with a prevalence of BA viruses between 80% and 100% in the last tree seasons.

Overall, the data presented in Fig. 2 indicate that the BA genotype has increased in prevalence in every country analyzed so far and has reached a point where almost 100% of the B viruses identified after 2005-2006 contain the 60-nt duplication in the G protein gene.

Phylogenetic analyses and global lineage distribution patterns. In order to study the phylogeography of the BA genotype during 10 years of circulation (1998 to 2008), a phylogenetic analysis of the C-terminal fragment (nt 652 to 982) of the G gene sequences with the duplication available to the latest revision (April 2009) was made. A total of 208 unique sequences were used to construct the phylogenetic tree shown in Fig. 3. Although the number of sequences has increased 4-fold with respect to our previous work, the majority of G sequences with the duplicated segment clustered in the same six branches (BA-I to BA-VI) previously described for the BA genotype (64), and the topology of the tree remains unchanged (Fig. 3A), conferring consistency to the analysis.

The first viruses with the 60-nt duplication that circulated in Madrid, Buenos Aires, and Belgium between 1998 and 2000 clustered in the BA-I branch, which prevailed over three epidemic seasons and disappeared in 2002. Despite the incorporation of new strains, all branches remained roughly unaltered in comparison with the phylogenetic analysis of our previous work (64). An exception to this was BA-IV. Viruses of this branch, which were a minority among group B viruses in our previous analysis (64), have spread worldwide. In addition, about 70% of all the BA sequences available now cluster in the BA-IV branch, including all the G sequences with the 60-nt duplication from samples collected after 2005. Sequences included in this branch were found to incorporate nucleotide changes over time, which split BA-IV into different sublineages: BA-IVa, BA-IVb, MAD-II, MAD-III, INDIA, and BRAZIL (Fig. 3B).

In general, a certain geographic clustering during a short period of time was observed for certain branches. For instance, temporary and geographic clustering was found in the sublineages MAD-II and MAD-III (Fig. 3B), which included sequences from viruses isolated only in Madrid in 2003-2004/2004-2005 and 2006-2007/2007-2008, respectively. Other significant examples of temporary and geographic clustering are constituted by the sequences from India and Brazil (Fig. 3B; see also Fig. S2 in the supplemental material). Thus, 79.4% of the sequences from India, over five epidemic seasons, clustered in the same BA-IV sublineage, named INDIA, and 93.9% of the viruses with the duplication that circulated in Brazil between 2005 and 2007 grouped in the BA-IV sublineage named BRAZIL. Therefore, the temporary and geographic clustering could range from 2 to 5 years. The large number of samples analyzed in Brazil rules out the possibility of a bias in clustering due to inadequate sampling. The sequences from Madrid that clustered in the MAD-IV branch of Fig. 1 were subsequently clustered in the BRAZIL and BA-III branches of the Fig. 3 tree and probably represent viruses imported from the Southern Hemisphere.

Multiple identical sequences were found in each region during the same epidemic season, and occasionally the same sequence was found in two or three successive epidemics (Fig. 3, sequences marked with asterisk). However, it was also observed that viruses of the same lineage may have circulated during the same season in very distant places. For instance, viruses of the sublineage BA-IVa were circulating in Spain, Belgium, and Canada in 2001-2002, or viruses of the branch BA-VI were isolated in Belgium, Canada, and Japan in the 2002-2003 season (Fig. 3A). Moreover, identical sequences

from viruses isolated in very distant regions were found circulating in successive epidemics, e.g., the strain QUE/85/03 appeared in Canada in 2003-2004, next in Buenos Aires in 2004 (BA/354/04), in Belgium in 2004-2005 and 2005-2006 (BE/12817/04), and finally in Brazil in 2006 (SP/1104/06).

Four sequences did not group in any branch (QUE/108/01, BR/140/03, BE/302/04, and MAD/5015/05), but through the amino acid sequence analysis and the reconstruction of the ancestor by the Bayesian method, the first two could be clustered in branch BA-IV and the last two in branch BA-II. These strains are shown as NC (not clustered) in Fig S3 in the supplemental material.

The alignment of partial G nucleotide sequences from samples of the BA genotype of Fig. 3 is shown in Fig. S3 in the supplemental material. Since the emergence of the first viruses of the BA genotype, with an identical copy of the duplicated segment, an accumulation over time of nucleotide substitutions, some of them resulting in amino acid changes, has been observed throughout the length of the fragment examined, including both copies of the duplication. Sixteen amino acid substitutions and three synonymous nucleotide changes were associated with the division of the BA genotype into the different lineages and are indicated at the origin of the main branches in the tree of Fig. 3A and B.

According to recent work, 8 of the 16 amino acid changes that defined the different lineages of the BA genotype were found to be subject to positive selection (summarized in Fig. 3, indicated with green rectangles), two of them (residues 267 and 270) within the duplication (7, 73).

Another important source of variation in the G protein of the BA genotype was the protein length polymorphism. By virtue of the use of three alternative stop codons described previously for group B viruses (41), three prevailing protein lengths of 312, 315, and 319 amino acids were observed (Fig. 3). Notably, the change 952CAA → UAA (Q313 → STOP) regenerated the first alternative stop codon found in group B viruses, shortening the polypeptide length to 312 amino acids. This change has been reported by Botosso et al. to be under positive selection (7) and was present in viruses of the most recent lineages of the BA-IV branch. In addition, several sequences showed less-frequent termination codons and other unusual changes (insertions and/or deletions) outside the duplicated segment that contributed to the protein length polymorphism. In summary, the sequences of the BA genotype showed a total of 11 different protein lengths, with a relatively good correlation between the protein length and the classification of sequences in different branches (see Fig. S3 in the supplemental material).

The prevalence and distribution of the different lineages shown in Fig. 3 are represented in Fig. 4 according to the year of isolation and place of origin. In general, multiple lineages cocirculated during the same epidemic, but in the majority of countries, one lineage prevailed over the rest. Furthermore, persistence of a given lineage was observed for more than one epidemic season in the same place. For instance, in Belgium the same pattern of circulating lineages was observed in two consecutive seasons (2001-2002 and 2002-2003), and viruses of the BA-VI branch were present in five consecutive seasons (2001-2002 to 2005-2006), although with different prevalences (Fig. 4). Similarly, the same sublineage (BA-IV INDIA) pre-

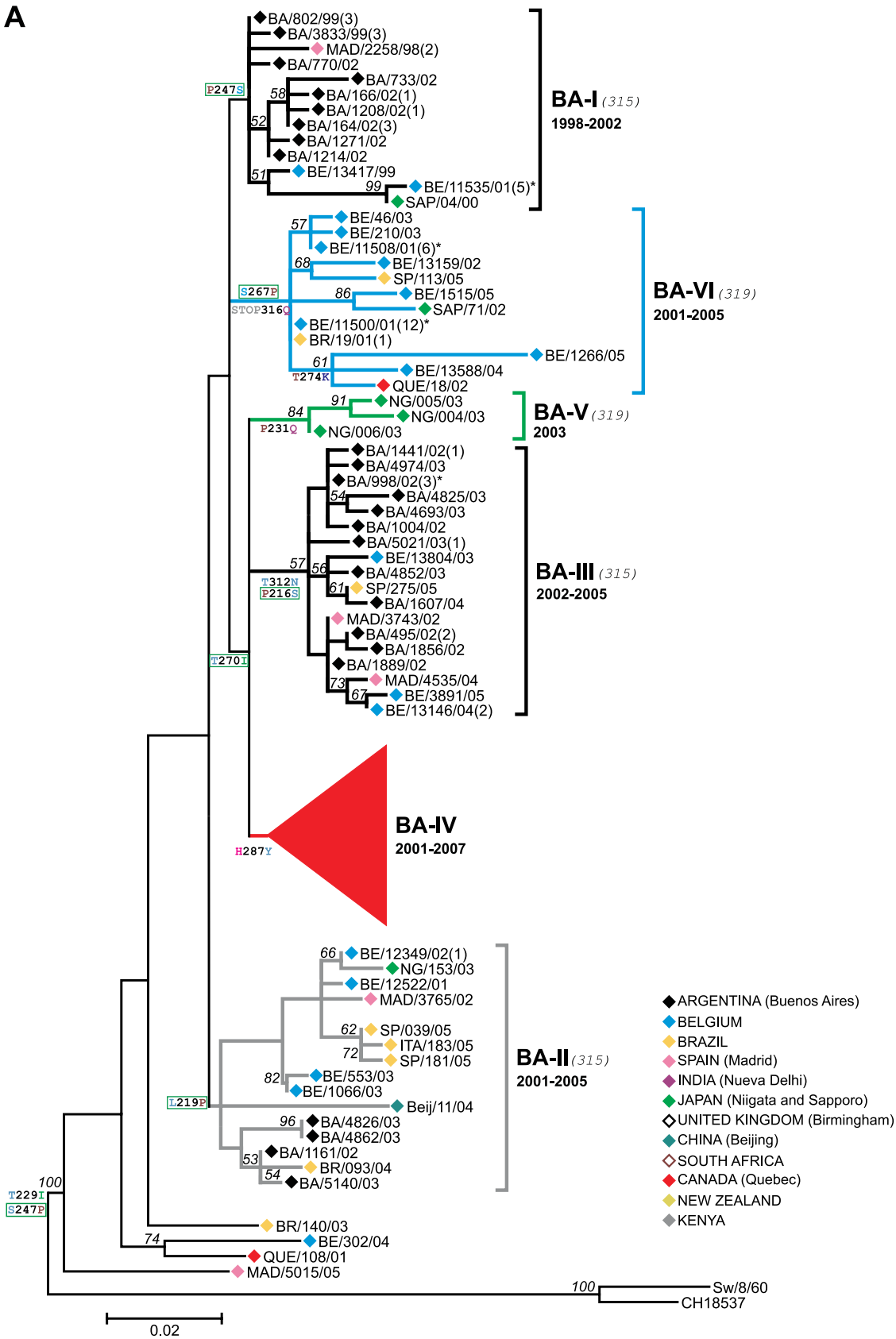
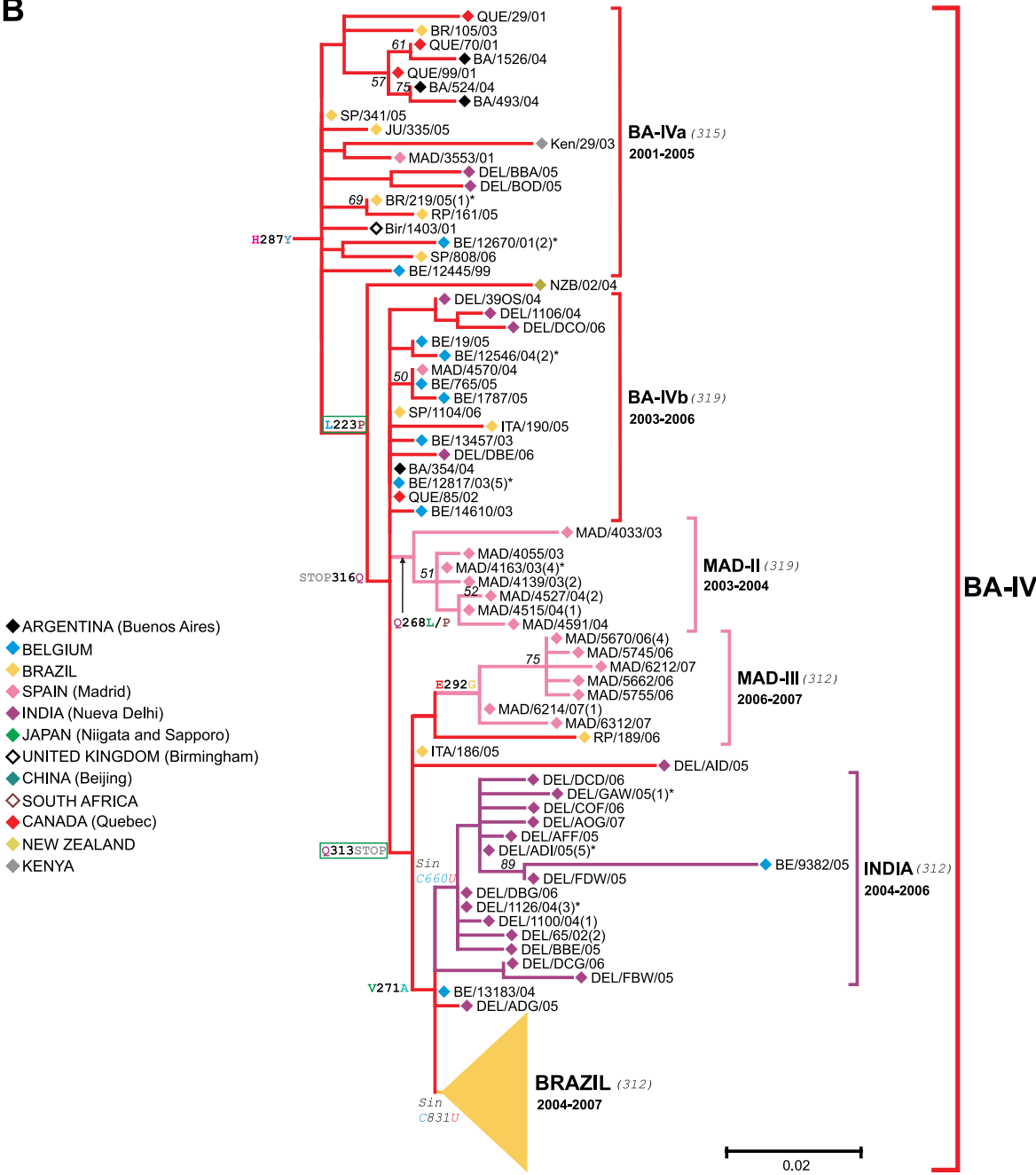
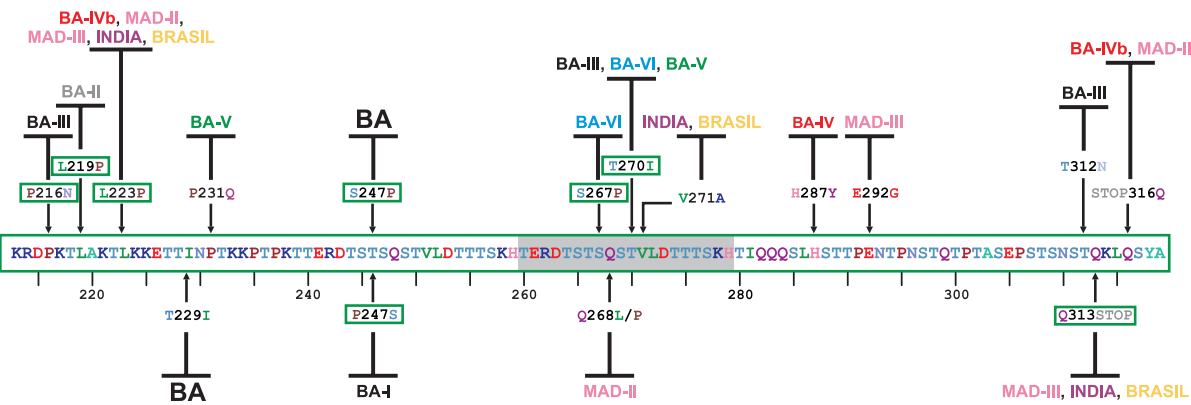


FIG. 3–Continued on following page

B



C



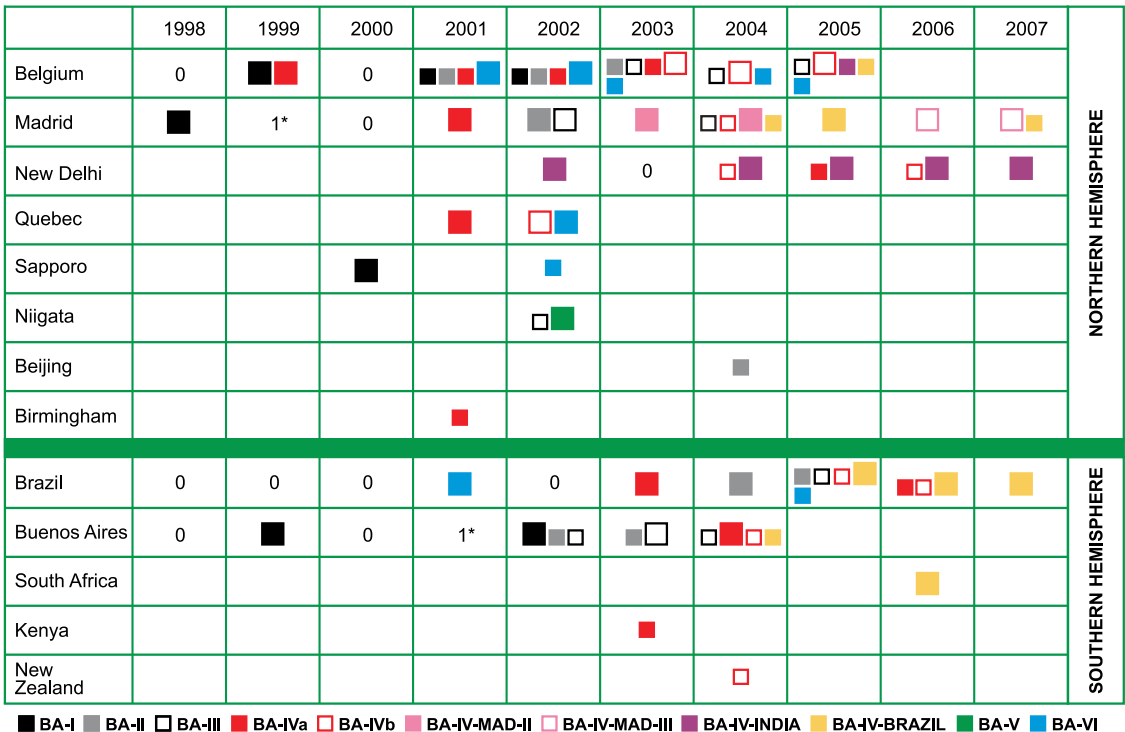


FIG. 4. Lineage distribution of the BA genotype. The lineages analyzed in Fig. 2 are represented in terms of year and place of isolation. Predominant branches or sublineages ($\geq 50\%$ of frequency over total sequences) are represented with larger boxes. The empty cells indicate an absence of epidemiologic data. Zero means that no samples with duplication were found. *, sample with duplication for which the sequence could not be obtained.

ailed in India for five seasons (2002-2003 and 2004-2005 to 2007-2008) or in Brazil (BA-IV BRAZIL) for three years (2005 to 2007). On average, we found that viruses of the same branch circulated in the same region between 2 and 4 years, prevailing during two to three epidemic seasons.

Evolutionary dynamics of the BA genotype. In order to estimate the rate of nucleotide substitutions and the time of the most recent common ancestor (TMRCA) and to reconstruct the demographic history of the BA genotype through 10 years of global circulation, we conducted a detailed Bayesian Markov chain Monte Carlo (MCMC) phylogenetic analysis (20, 21) combined with the Bayesian skyline plot method (22) with the 208 partial G sequences used in Fig. 3.

The rate was 4.7×10^{-3} nucleotide substitutions/site/year

(HPD, 95%: 3.7×10^{-3} to 5.7×10^{-3} nucleotide substitutions/site/year), and the TMRCA was 1997 (HPD, 95%: 1994 to 1999). Despite including in the analysis only the C-terminal 330 nt of a large number of sequences, the rate and TMRCA values are close to those previously reported for a limited number of complete G sequences (64). The rate of substitutions was greater than that reported for the complete gene, as might be expected for a highly variable G region.

The maximum clade credibility tree and the Bayesian skyline plot depicting the demographic history of the BA genotype are represented in the same time scale in Fig. 5. The demographic history is represented as the changes in the effective population of viral infections over time.

Since the first reported viruses, the BA genotype has showed

FIG. 3. Phylogenetic tree of the BA genotype, constructed using the maximum-likelihood method. (A) A total of 208 unique partial G sequences (nucleotides 652 to 982) with the 60-nt duplication from samples isolated in Buenos Aires (BA), Belgium (BE), Brazil (SP, RP, BR, JU, and ITA), Madrid (MAD), India (DEL), Niigata (NG), Sapporo (SAP), Birmingham (Bir), Beijing (Beij), Quebec (QUE), Kenya (Ken), South Africa (SA), and New Zealand (NZB) were used to construct the phylogram by the maximum-likelihood method (see Methods). The samples are marked with diamonds color coded by country. The bar represents 0.02 nucleotide substitutions per site, and the tree was rooted with the CH18537 and Sw/8/60 strains. Numbers at the internal nodes represent the bootstrap probabilities (1,000 replicates). Only bootstrap values > 50 are shown. Names of viruses refer to place/number/year of isolation. Branches and sublineages are indicated at right by brackets, together with the period of most frequent isolation and the G protein mean length (in amino acids) of sequences in each branch (between parentheses). The number of sequences identical to those shown in the figure is indicated between parentheses at the right of the sample name, and the asterisk indicates sequences identified in more than one season. The amino acid substitutions that originate the main branches are indicated at the left of the nodes, and the residues with evidence of positive selection are framed by green rectangles (7, 73). (B) Expanded BA-IV branch. Synonymous nucleotide substitutions implicated in the origin of certain sublineages are indicated at the left of the nodes with *Sin*. (C) Scheme of the third C-terminal region of the BA G protein. The amino acid changes associated with division of the BA genotype in different lineages are indicated by arrows. The residues with evidence of positive selection are framed by green rectangles (7, 73). The amino acid sequence is from strain BA/802/99, modified to show the alternative termination codons. The 20 duplicated amino acids are shaded in gray.

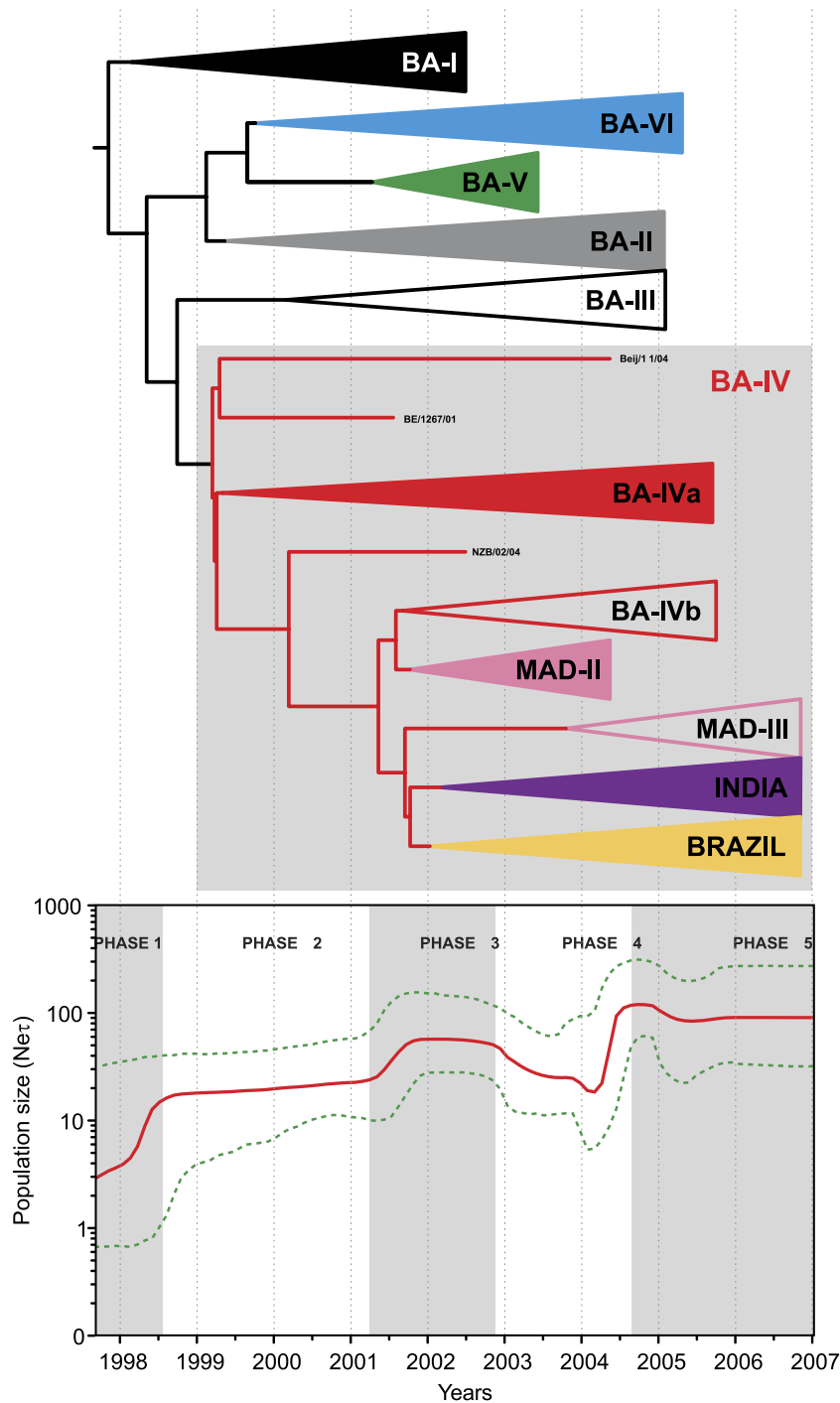


FIG. 5. Genealogies and corresponding Bayesian skyline plot showing the demographic history of BA sequences, drawn on the same time scale. The 208 unique partial G sequences with 60-nt duplication represented in Fig. 2 were used to infer the maximum clade credibility (MCC) tree and the skyline plot through the Bayesian Markov chain Monte Carlo analysis implemented in the BEAST software program (21, 22). The y axes of the skyline plot (lower panel) represent the population size, which is equal to the product of the effective population size (N_e) and the generation length in years (τ). The red line represents the median estimate, and the areas between the green dotted lines show the 95% HPD limits. The MCC tree (upper panel) is represented on the same time scale as the skyline plot. Branches were collapsed for clarity, and the tree is rooted with the CH18537 and Sw/8/60 strains (not shown).

a demographic expansion, with fluctuations that could be divided into five phases (Fig. 5). A short initial invasion (phase 1) was followed by a moderate expansion (phase 2), which then coincided with the first BA viruses clustered in the BA-I

branch and their moderate dissemination. Subsequently, from the middle of 2001 to the end of 2002, new growth in the effective population was observed (phase 3), which corresponds with the appearance and fast dissemination of new BA

lineages and the extinction of the BA-I branch. At a later time, the viral effective population showed a depression, followed by a rapid increase toward 2005 (phase 4) and a new stabilization (phase 5). The constriction and subsequent expansion in the effective population observed in phase 4 are characteristic of bottleneck effects (22, 39) and coincide with the predominance of the BA genotype, particularly branch BA-IV, relative to the rest of the group B viruses. Although the confidence limits are ample and they are not statistically significant, these data, along with the results obtained in the phylogenetic analysis (Fig. 3) and the alignment of sequences (see Fig. S3 in the supplemental material), suggest that the small bottleneck of phase 4 could be related to the appearance of viruses of branch BA-IV with the substitution Q313 → STOP, which has been proposed to be under positive selection (7).

DISCUSSION

The unique duplication of 60 nt within the G protein gene of HRSV, first detected 10 years ago in viral isolates, has allowed us to essentially turn the evolutionary clock back to zero in order to better understand the natural history of this virus during propagation and dissemination in its natural host. Previous work provided strong evidence for a common ancestry of all viruses with this duplicated segment and showed that all sequences with this change clustered in a new genotype, named BA (64).

In Spain, as in the rest of the world, HRSV infections are the main cause of severe respiratory illness in the pediatric population (5, 29, 66) and in vulnerable adults (8, 38, 42). Although respiratory disease related to HRSV is clinically well characterized (2, 9, 10) and the risk factors for hospitalization associated with infection of this virus have been established (25, 58), only two previous studies with Spanish samples provided limited information about circulating genotypes (27, 41). The present study provides new epidemiological and phylogenetic information about HRSV circulation in Madrid during 12 consecutive epidemic seasons (1996-1997 to 2007-2008).

The first viruses with the duplication were found in Madrid in 1998, 6 months before the viruses reported from Buenos Aires in June 1999 (63), and therefore represent the earliest BA viruses described to this date (Table 1). The phylogenetic tree of Fig. 1 represents the circulation of BA viruses in Madrid during 10 epidemic seasons. Their G sequences clustered in four different branches (MAD-I to MAD-IV), which illustrate the local replacement of prevalent genotypes with time. In Madrid it appears that a given genotype typically remained prevalent over a 2-year period before replacement occurred.

Although viruses containing duplications or deletions of 1 or 2 codons in the G protein sequence have been described previously (6, 27, 43, 44, 73, 74), these changes appeared only sporadically during certain epidemics. Thus, it is remarkable that viruses containing the most extensive duplication described to date (60 nt) have been circulating worldwide for at least 10 consecutive years.

The epidemiological data from Madrid examined in this work, together with the data in recent publications (1, 7, 48, 55, 57, 68, 75), indicated that BA viruses have gradually replaced the rest of the group B genotypes (Fig. 2). The success of the BA genotype relative to the rest of the group B viruses suggests

that the 60-nt duplication in the G protein gene provided some advantage over related viruses lacking this duplication. However, this advantage has not affected the alternating prevalences of A and B antigenic groups during epidemics. As with the vast majority of HRSV strain variability studies, our epidemiologic analysis is limited to samples obtained from hospitalized infants. However, several comparative studies between samples from hospitalized babies and community patients found similar strain variability in the two type of samples (46, 56, 65, 72).

The mechanism by which this duplication provides a selective advantage for the BA genotype over the rest of the B viruses is unknown; however, it is possible that the additional residues modified significantly the antigenic characteristics of the G protein, allowing escape from neutralization by antibodies present in the human population as a result of previous infections. An additional possibility is that the duplication has affected G protein-mediated attachment of the virus to the host cell and thereby enhanced virus fitness. However, since the entire genome sequence of a virus of the BA genotype remains to be determined, it cannot be excluded that viruses of the BA genotype contain additional mutations, besides the duplication, that may increase their fitness relative to that of other group B viruses.

The phylogenetic tree in Fig. 3 represents the global spread of the BA genotype during 10 years. The six branches (BA-I to BA-VI) previously identified contain the majority of viruses with the duplication (64), but recently the branch BA-IV has undergone a great expansion relative to other branches and now holds 70% of the sequences analyzed and all BA sequences identified after 2005. This observation suggests that viruses of this branch, particularly the recent predominant lineages MAD-III, INDIA, and BRASIL (Fig. 3B), have been more successful than the remainder of the BA genotype, presumably due to some selective advantage.

The locations where BA viruses have been identified, as well as the different lineages and the year of isolation, are illustrated in the map of Fig. 6. Although it has been previously reported that HRSV has a low tendency for geographical association (12), two different situations were observed for the BA genotype, one where certain lineages circulated in restricted areas during at least 2 years (BA-III, BA-V, MAD-II, MAD-III, INDIA, and BRAZIL) and another in which a new variant of the virus reached very distant places in a short period of time. These two situations are probably the result of a complex interaction between a high substitution rate and the antigenic drift imposed by the immunological status of the population. It is expected that HRSV, like many other viruses, is permanently engaged in an arms race with the immune system of the host and that the ability to escape detection by acquired immunity plays a central role in the progressive evolution of HRSV.

For HRSV group A, it has been documented that amino acid changes in the C-terminal third of the G protein are often associated with the loss of antigenic epitopes in both natural strains and antibody escape mutants (4, 11, 15, 24, 27, 36, 40, 54). These changes coincide to some extent with sites described as being under positive selection (7, 43, 71, 74). Although less is known about group B virus epitopes, it is probable that the sites associated with the division of the BA genotype in lin-

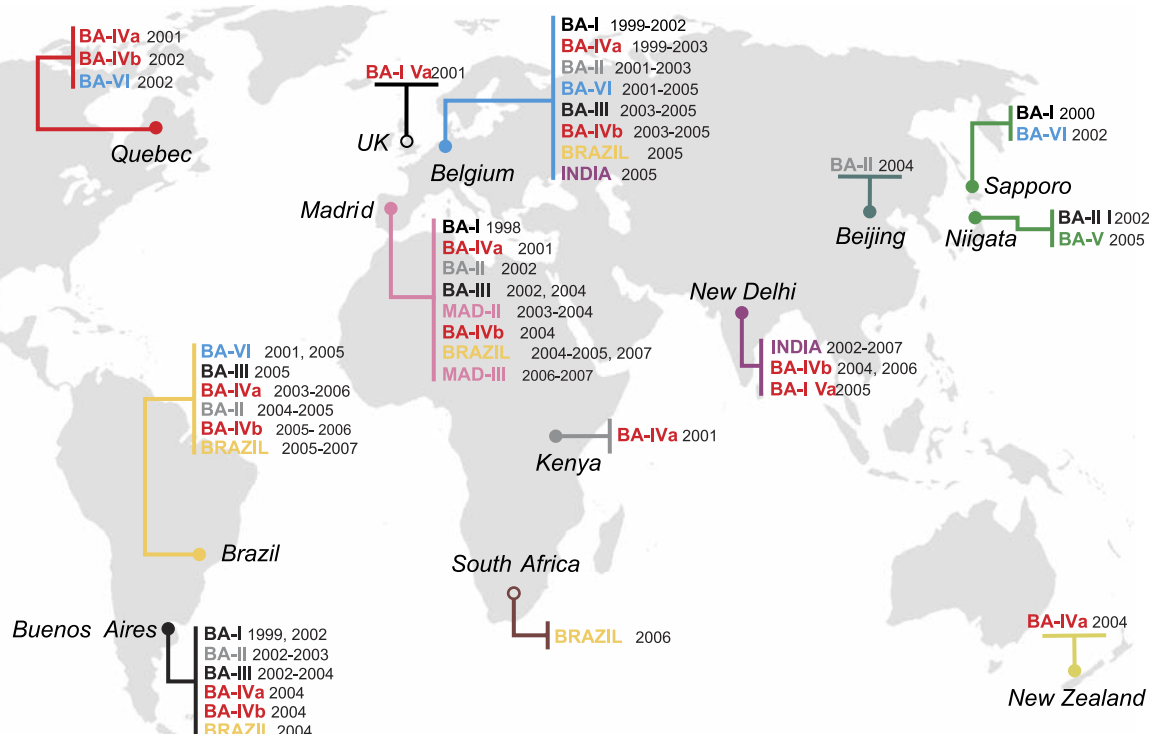


FIG. 6. Global distribution of BA genotype sequences. The location from which HRSV G sequences with the duplicated segment have been reported until December 2008 are shown in the map (color coded as described for Fig. 3), indicating BA lineages and epidemics, separated by “–” for consecutive seasons and by “,” for nonconsecutive seasons. The year of isolation corresponds to the first year of an epidemic season, i.e., 2001 for the 2001-2002 epidemic season.

eages and sublineages and those found to be under positive selection (Fig. 3C) represent antigenic changes that contribute to the evasion of the immune response by the population.

An interesting observation is the high degree of G protein length polymorphism found among BA viruses, which showed a total of 11 different G protein lengths (see Fig. S3 in the supplemental material). Variation in G protein length has been associated with selection of certain antibody escape mutants (54).

The demographic history of the BA genotype represented in Fig. 5 displays a general expansion of the viral population, with fluctuations imposed by the selective forces that operate in HRSV evolution. It is important to note the decrease and subsequent increase in the viral population, observed in 2004, characteristic of a bottleneck effect. This point in the demography coincides with the replacement of all circulating B genotypes by the BA-IV branch, which predominated over all other BA lineages. Despite the capacity for a rapid global spread of the BA virus, probably as a consequence of the immunologically naive conditions of the population, these viruses did not dominate HRSV group B until a few years later, when branch BA-IV became predominant. This observation suggests that other changes besides the 60-nt duplication might have been required for optimal adaptation of the new virus for replication in its natural host, e.g., the shortening of the protein length found in most recent lineages of the BA-IV branch.

Future studies will determine whether this new BA genotype continues to dominate the other group B viruses and becomes the new antigenic HRSV group B or BA viruses are lost from

the virus pool as a consequence of changes in the herd immunity.

ACKNOWLEDGMENTS

We are most grateful to Monica Galiano for constructive comments on an earlier version of the work and to Keith Chappell for critical reading of the manuscript and language contribution.

This work was supported by grant SAF2009-11632 from Ministerio de Ciencia e Innovación. J.A.M. participates in the VIRHOST consortium funded by Comunidad de Madrid.

REFERENCES

1. Agrawal, A. S., M. Sarkar, S. Ghosh, M. Chawla-Sarkar, N. Chakraborty, M. Basak, and T. N. Naik. 2009. Prevalence of respiratory syncytial virus group B genotype BA-IV strains among children with acute respiratory tract infection in Kolkata, Eastern India. *J. Clin. Virol.* **45**:358–361.
2. Alonso, A., J. M. Andres, J. R. Garmendia, I. Diez, J. M. Gil, and J. Ardura. 2007. Bronchiolitis due to respiratory syncytial virus in hospitalized children: a study of seasonal rhythm. *Acta Paediatr.* **96**:731–735.
3. Anderson, L. J., R. M. Hendry, L. T. Pierik, C. Tsou, and K. McIntosh. 1991. Multicenter study of strains of respiratory syncytial virus. *J. Infect. Dis.* **163**:687–692.
4. Anderson, L. J., J. C. Hierholzer, C. Tsou, R. M. Hendry, B. F. Fernie, Y. Stone, and K. McIntosh. 1985. Antigenic characterization of respiratory syncytial virus strains with monoclonal antibodies. *J. Infect. Dis.* **151**:626–633.
5. Artiles-Campelo, F., M. C. Perez-Gonzalez, A. Caballero-Hidalgo, and M. J. Pena-Lopez. 2006. Etiology of acute viral respiratory tract infections in children from Gran Canaria, the Canary Islands (Spain). *Enferm. Infecc. Microbiol. Clin.* **24**:556–561. (In Spanish.)
6. Blanc, A., A. Delfraro, S. Frabasile, and J. Arbiza. 2005. Genotypes of respiratory syncytial virus group B identified in Uruguay. *Arch. Virol.* **150**:603–609.
7. Botosso, V. F., P. M. Zanotto, M. Ueda, E. Arruda, A. E. Gilio, S. E. Vieira, K. E. Stewien, T. C. Peret, L. F. Jamal, M. I. Pardini, J. R. Pinho, E. Massad, O. A. Sant'anna, E. C. Holmes, and E. L. Durigon. 2009. Positive selection

- results in frequent reversible amino acid replacements in the G protein gene of human respiratory syncytial virus. *PLoS Pathog.* 5:e1000254.
8. **Bueno Campaña, M., C. Calvo Rey, M. C. Vazquez Alvarez, E. Parra Cuadrado, C. Molina Amores, G. Rodrigo García, F. Echavarrí Olavarría, J. Valverde Cánovas, and I. Casas Flecha.** 2008. Viral respiratory tract infections in the first six months of life. *An. Pediatr. (Barc.)* 69:400–405. (In Spanish.)
 9. **Calvo, C., M. L. Garcia-Garcia, C. Blanco, M. C. Vazquez, M. E. Frias, P. Perez-Brena, and I. Casas.** 2008. Multiple simultaneous viral infections in infants with acute respiratory tract infections in Spain. *J. Clin. Virol.* 42:268–272.
 10. **Calvo, C., M. L. Garcia-Garcia, F. Pozo, O. Carvajal, P. Perez-Brena, and I. Casas.** 2008. Clinical characteristics of human bocavirus infections compared with other respiratory viruses in Spanish children. *Pediatr. Infect. Dis. J.* 27:677–680.
 11. **Cane, P. A.** 1997. Analysis of linear epitopes recognised by the primary human antibody response to a variable region of the attachment (G) protein of respiratory syncytial virus. *J. Med. Virol.* 51:297–304.
 12. **Cane, P. A.** 2007. Molecular epidemiology and evolution of RSV, p. 89–113. *In* P. A. Cane (ed.), *Perspectives in medical virology*, vol. 14. Respiratory syncytial virus. Elsevier, Salisbury, United Kingdom.
 13. **Cane, P. A., D. A. Matthews, and C. R. Pringle.** 1992. Analysis of relatedness of subgroup A respiratory syncytial viruses isolated worldwide. *Virus Res.* 25:15–22.
 14. **Cane, P. A., D. A. Matthews, and C. R. Pringle.** 1994. Analysis of respiratory syncytial virus strain variation in successive epidemics in one city. *J. Clin. Microbiol.* 32:1–4.
 15. **Cane, P. A., and C. R. Pringle.** 1995. Evolution of subgroup A respiratory syncytial virus: evidence for progressive accumulation of amino acid changes in the attachment protein. *J. Virol.* 69:2918–2925.
 16. **Coates, H. V., D. W. Alling, and R. M. Chanock.** 1966. An antigenic analysis of respiratory syncytial virus isolates by a plaque reduction neutralization test. *Am. J. Epidemiol.* 83:299–313.
 17. **Coiras, M. T., P. Perez-Brena, M. L. Garcia, and I. Casas.** 2003. Simultaneous detection of influenza A, B, and C viruses, respiratory syncytial virus, and adenoviruses in clinical samples by multiplex reverse transcription nested-PCR assay. *J. Med. Virol.* 69:132–144.
 18. **Cristina, J., J. A. Lopez, C. Albo, B. Garcia-Barreno, J. Garcia, J. A. Melero, and A. Portela.** 1990. Analysis of genetic variability in human respiratory syncytial virus by the RNase A mismatch cleavage method: subtype divergence and heterogeneity. *Virology* 174:126–134.
 19. **Drummond, A. J., S. Y. Ho, M. J. Phillips, and A. Rambaut.** 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4:e88.
 20. **Drummond, A. J., G. K. Nicholls, A. G. Rodrigo, and W. Solomon.** 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 161:1307–1320.
 21. **Drummond, A. J., and A. Rambaut.** 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC. Evol. Biol.* 7:214.
 22. **Drummond, A. J., A. Rambaut, B. Shapiro, and O. G. Pybus.** 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* 22:1185–1192.
 23. **Falsey, A. R., P. A. Hennessey, M. A. Formica, C. Cox, and E. E. Walsh.** 2005. Respiratory syncytial virus infection in elderly and high-risk adults. *N. Engl. J. Med.* 352:1749–1759.
 24. **Falsey, A. R., and E. E. Walsh.** 2000. Respiratory syncytial virus infection in adults. *Clin. Microbiol. Rev.* 13:371–384.
 25. **Figueras, A. J., S. J. Lopez, L. C. Medrano, F. R. Bermudez-Canete, P. L. Fernandez, P. A. Bonillo, E. G. Perez-Yarza, H. C. az de, and T. J. Sanchez de.** 2008. Spanish multidisciplinary consensus on the prevention of respiratory syncytial virus infection in children. *An. Pediatr. (Barc.)* 69:63–71. (In Spanish.)
 26. **Galiano, M. C., C. Palomo, C. M. Videla, J. Arbiza, J. A. Melero, and G. Carballal.** 2005. Genetic and antigenic variability of human respiratory syncytial virus (groups a and b) isolated over seven consecutive seasons in Argentina (1995 to 2001). *J. Clin. Microbiol.* 43:2266–2273.
 27. **Garcia, O., M. Martin, J. Dopazo, J. Arbiza, S. Frabasile, J. Russi, M. Hortal, P. Perez-Brena, I. Martinez, B. Garcia-Barreno, and J. A. Melero.** 1994. Evolutionary pattern of human respiratory syncytial virus (subgroup A): cocirculating lineages and correlation of genetic and antigenic changes in the G glycoprotein. *J. Virol.* 68:5448–5459.
 28. **Garcia-Barreno, B., C. Palomo, C. Penas, T. Delgado, P. Perez-Brena, and J. A. Melero.** 1989. Marked differences in the antigenic structure of human respiratory syncytial virus F and G glycoproteins. *J. Virol.* 63:925–932.
 29. **Garcia Garcia, M. L., G. M. Ordobas, R. C. Calvo, A. M. Gonzalez, R. J. Aguilar, S. A. Arregui, and B. P. Perez.** 2001. Viral infection of the lower respiratory tract in hospitalized infants: etiology, clinical features and risk factors. *An. Esp. Pediatr.* 55:101–107. (In Spanish.)
 30. **Glezen, P., and F. W. Denny.** 1973. Epidemiology of acute lower respiratory disease in children. *N. Engl. J. Med.* 288:498–505.
 31. **Glezen, W. P., L. H. Taber, A. L. Frank, and J. A. Kasel.** 1986. Risk of primary infection and reinfection with respiratory syncytial virus. *Am. J. Dis. Child.* 140:543–546.
 32. **Hall, C. B., E. E. Walsh, K. C. Schnabel, C. E. Long, K. M. McConnochie, S. W. Hildreth, and L. J. Anderson.** 1990. Occurrence of groups A and B of respiratory syncytial virus over 15 years: associated epidemiologic and clinical characteristics in hospitalized and ambulatory children. *J. Infect. Dis.* 162:1283–1290.
 33. **Hall, T. A.** 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis. North Carolina State University, Raleigh, NC.
 34. **Henderson, F. W., A. M. Collier, W. A. Clyde, Jr., and F. W. Denny.** 1979. Respiratory-syncytial-virus infections, reinfections and immunity. A prospective, longitudinal study in young children. *N. Engl. J. Med.* 300:530–534.
 35. **Holberg, C. J., A. L. Wright, F. D. Martinez, C. G. Ray, L. M. Taussig, and M. D. Lebowitz.** 1991. Risk factors for respiratory syncytial virus-associated lower respiratory illnesses in the first year of life. *Am. J. Epidemiol.* 133:1135–1151.
 36. **Johnson, P. R., Jr., R. A. Olmsted, G. A. Prince, B. R. Murphy, D. W. Alling, E. E. Walsh, and P. L. Collins.** 1987. Antigenic relatedness between glycoproteins of human respiratory syncytial virus subgroups A and B: evaluation of the contributions of F and G glycoproteins to immunity. *J. Virol.* 61:3163–3166.
 37. **Johnson, P. R., M. K. Spriggs, R. A. Olmsted, and P. L. Collins.** 1987. The G glycoprotein of human respiratory syncytial viruses of subgroups A and B: extensive sequence divergence between antigenically related proteins. *Proc. Natl. Acad. Sci. U. S. A.* 84:5625–5629.
 38. **Lopez-Medrano, F., J. M. Aguado, M. Lizasoain, D. Folgueira, R. S. Juan, C. az-Pedroche, C. Lumbreras, J. M. Morales, J. F. Delgado, and E. Moreno-Gonzalez.** 2007. Clinical implications of respiratory virus infections in solid organ transplant recipients: a prospective study. *Transplantation* 84:851–856.
 39. **Manrubia, S. C., C. Escarmis, E. Domingo, and E. Lazaro.** 2005. High mutation rates, bottlenecks, and robustness of RNA viral quasispecies. *Gene* 347:273–282.
 40. **Martinez, I., J. Dopazo, and J. A. Melero.** 1997. Antigenic structure of the human respiratory syncytial virus G glycoprotein and relevance of hypermutation events for the generation of antigenic variants. *J. Gen. Virol.* 78(Pt. 10):2419–2429.
 41. **Martinez, I., O. Valdes, A. Delfraro, J. Arbiza, J. Russi, and J. A. Melero.** 1999. Evolutionary pattern of the G glycoprotein of human respiratory syncytial viruses from antigenic group B: the use of alternative termination codons and lineage diversification. *J. Gen. Virol.* 80(Pt. 1):125–130.
 42. **Martino, R., R. P. Porras, N. Rabella, J. V. Williams, E. Ramila, N. Margall, R. Labeaga, J. E. Crowe, Jr., P. Coll, and J. Sierra.** 2005. Prospective study of the incidence, clinical features, and outcome of symptomatic upper and lower respiratory tract infections by respiratory viruses in adult recipients of hematopoietic stem cell transplants for hematologic malignancies. *Biol. Blood Marrow Transplant.* 11:781–796.
 43. **Melero, J. A., B. Garcia-Barreno, I. Martinez, C. R. Pringle, and P. A. Cane.** 1997. Antigenic structure, evolution and immunobiology of human respiratory syncytial virus attachment (G) protein. *J. Gen. Virol.* 78(Pt. 10):2411–2418.
 44. **Moura, F. E., A. Blanc, S. Frabasile, A. Delfraro, M. J. de Sierra, L. Tome, E. A. Ramos, M. M. Siqueira, and J. Arbiza.** 2004. Genetic diversity of respiratory syncytial virus isolated during an epidemic period from children of northeastern Brazil. *J. Med. Virol.* 74:156–160.
 45. **Mufson, M. A., C. Orvell, B. Rafnar, and E. Norrby.** 1985. Two distinct subtypes of human respiratory syncytial virus. *J. Gen. Virol.* 66(Pt. 10):2111–2124.
 46. **Nokes, D. J., E. A. Okiro, M. Ngama, L. J. White, R. Ochola, P. D. Scott, P. A. Cane, and G. F. Medley.** 2004. Respiratory syncytial virus epidemiology in a birth cohort from Kilifi district, Kenya: infection during the first year of life. *J. Infect. Dis.* 190:1828–1832.
 47. **Page, R. D.** 1996. TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* 12:357–358.
 48. **Parveen, S., S. Broor, S. K. Kapoor, K. Fowler, and W. M. Sullender.** 2006. Genetic diversity among respiratory syncytial viruses that have caused repeated infections in children from rural India. *J. Med. Virol.* 78:659–665.
 49. **Peret, T. C., C. B. Hall, G. W. Hammond, P. A. Piedra, G. A. Storch, W. M. Sullender, C. Tsou, and L. J. Anderson.** 2000. Circulation patterns of group A and B human respiratory syncytial virus genotypes in 5 communities in North America. *J. Infect. Dis.* 181:1891–1896.
 50. **Peret, T. C., C. B. Hall, K. C. Schnabel, J. A. Golub, and L. J. Anderson.** 1998. Circulation patterns of genetically distinct group A and B strains of human respiratory syncytial virus in a community. *J. Gen. Virol.* 79(Pt. 9):2221–2229.
 51. **Posada, D., and K. A. Crandall.** 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14:817–818.
 52. **Roberts, S. R., D. Lichtenstein, L. A. Ball, and G. W. Wertz.** 1994. The membrane-associated and secreted forms of the respiratory syncytial virus attachment glycoprotein G are synthesized from alternative initiation codons. *J. Virol.* 68:4538–4546.
 53. **Rozas, J., J. C. Sanchez-DelBarrio, X. Messeguer, and R. Rozas.** 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497.
 54. **Rueda, P., T. Delgado, A. Portela, J. A. Melero, and B. Garcia-Barreno.**

1991. Premature stop codons in the G glycoprotein of human respiratory syncytial viruses resistant to neutralization by monoclonal antibodies. *J. Virol.* **65**:3374–3378.
55. Sato, M., R. Saito, T. Sakai, Y. Sano, M. Nishikawa, A. Sasaki, Y. Shobugawa, F. Gejyo, and H. Suzuki. 2005. Molecular epidemiology of respiratory syncytial virus infections among children with acute respiratory symptoms in a community over three seasons. *J. Clin. Microbiol.* **43**:36–40.
56. Scott, P. D., R. Ochola, M. Ngama, E. A. Okiro, D. J. Nokes, G. F. Medley, and P. A. Cane. 2004. Molecular epidemiology of respiratory syncytial virus in Kilifi district, Kenya. *J. Med. Virol.* **74**:344–354.
57. Shobugawa, Y., R. Saito, Y. Sano, H. Zaraket, Y. Suzuki, A. Kumaki, I. Daput, T. Oguma, M. Yamaguchi, and H. Suzuki. 2009. Emerging genotypes of human respiratory syncytial virus subgroup A among patients in Japan. *J. Clin. Microbiol.* **47**:2475–2482.
58. Simoes, E. A., X. Carbonell-Estrany, J. R. Fullarton, J. G. Liese, J. Figueras-Aloy, G. Doering, and J. Guzman. 2008. A predictive model for respiratory syncytial virus (RSV) hospitalisation of premature infants born at 33–35 weeks of gestational age, based on data from the Spanish FLIP Study. *Respir. Res.* **9**:78.
59. Sullender, W. M. 2000. Respiratory syncytial virus genetic and antigenic diversity. *Clin. Microbiol. Rev.* **13**:1–15, table.
60. Swofford, D. L. 1998. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, MA.
61. Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**:1596–1599.
62. Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**:4876–4882.
63. Trento, A., M. Galiano, C. Videla, G. Carballal, B. Garcia-Barreno, J. A. Melero, and C. Palomo. 2003. Major changes in the G protein of human respiratory syncytial virus isolates introduced by a duplication of 60 nucleotides. *J. Gen. Virol.* **84**:3115–3120.
64. Trento, A., M. Viegas, M. Galiano, C. Videla, G. Carballal, A. S. Mistchenko, and J. A. Melero. 2006. Natural history of human respiratory syncytial virus inferred from phylogenetic analysis of the attachment (G) glycoprotein with a 60-nucleotide duplication. *J. Virol.* **80**:975–984.
65. Venter, M., M. Collinson, and B. D. Schoub. 2002. Molecular epidemiological analysis of community circulating respiratory syncytial virus in rural South Africa: comparison of viruses and genotypes responsible for different disease manifestations. *J. Med. Virol.* **68**:452–461.
66. Vicente, D., M. Montes, G. Cilla, E. G. Perez-Yarza, and E. Perez-Trallero. 2003. Hospitalization for respiratory syncytial virus in the paediatric population in Spain. *Epidemiol. Infect.* **131**:867–872.
67. Viegas, M., and A. S. Mistchenko. 2005. Molecular epidemiology of human respiratory syncytial virus subgroup A over a six-year period (1999–2004) in Argentina. *J. Med. Virol.* **77**:302–310.
68. Visser, A., S. Delpont, and M. Venter. 2008. Molecular epidemiological analysis of a nosocomial outbreak of respiratory syncytial virus associated pneumonia in a kangaroo mother care unit in South Africa. *J. Med. Virol.* **80**:724–732.
69. Wertz, G. W., P. L. Collins, Y. Huang, C. Gruber, S. Levine, and L. A. Ball. 1985. Nucleotide sequence of the G protein gene of human respiratory syncytial virus reveals an unusual type of viral membrane protein. *Proc. Natl. Acad. Sci. U. S. A.* **82**:4075–4079.
70. WHO. 2009. Initiative for Vaccine Research (IVR). Respiratory syncytial virus and parainfluenza viruses. Disease burden. WHO, Geneva, Switzerland. http://www.who.int/vaccine_research/diseases/ari/en/index2.html.
71. Woelk, C. H., and E. C. Holmes. 2001. Variable immune-driven natural selection in the attachment (G) glycoprotein of respiratory syncytial virus (RSV). *J. Mol. Evol.* **52**:182–192.
72. Zambon, M. C., J. D. Stockton, J. P. Clewley, and D. M. Fleming. 2001. Contribution of influenza and respiratory syncytial virus to community cases of influenza-like illness: an observational study. *Lancet* **358**:1410–1416.
73. Zlateva, K. T., P. Lemey, E. Moes, A. M. Vandamme, and M. Van Ranst. 2005. Genetic variability and molecular evolution of the human respiratory syncytial virus subgroup B attachment G protein. *J. Virol.* **79**:9157–9167.
74. Zlateva, K. T., P. Lemey, A. M. Vandamme, and M. Van Ranst. 2004. Molecular evolution and circulation patterns of human respiratory syncytial virus subgroup A: positively selected sites in the attachment G glycoprotein. *J. Virol.* **78**:4675–4683.
75. Zlateva, K. T., L. Vijgen, N. Dekeersmaecker, C. Naranjo, and M. Van Ranst. 2007. Subgroup prevalence and genotype circulation patterns of human respiratory syncytial virus in Belgium during ten successive epidemic seasons. *J. Clin. Microbiol.* **45**:3022–3030.